2017-07-01

Alternative infill strategies for expensive multi-objective optimisation

Rahat, A

http://hdl.handle.net/10026.1/12584

10.1145/3071178.3071276

GECCO 2017 - Proceedings of the 2017 Genetic and Evolutionary Computation Conference

Alternative Infill Strategies for Expensive Multi-Objective Optimisation

Alma A. M. Rahat*
University of Exeter
United Kingdom
A.A.M.Rahat@exeter.ac.uk

Richard M. Everson
University of Exeter
United Kingdom
R.M.Everson@exeter.ac.uk

Jonathan E. Fieldsend
University of Exeter
United Kingdom
J.E.Fieldsend@exeter.ac.uk

ABSTRACT

Many multi-objective optimisation problems incorporate computationally or financially expensive objective functions. State-of-the-art algorithms therefore construct surrogate model(s) of the parameter space to objective functions mapping to guide the choice of the next solution to expensively evaluate. Starting from an initial set of solutions, an infill criterion — a surrogate-based indicator of quality — is extremised to determine which solution to evaluate next, until the budget of expensive evaluations is exhausted. Many successful infill criteria are dependent on multi-dimensional integration, which may result in infill criteria that are themselves impractically expensive. We propose a computationally cheap infill criterion based on the minimum probability of improvement over the estimated Pareto set. We also present a range of set-based scalarisation methods modelling hypervolume contribution, dominance ratio and distance measures. These permit the use of straightforward expected improvement as a cheap infill criterion. We investigated the performance of these novel strategies on standard multi-objective test problems, and compared them with the popular SMS-EGO and ParEGO methods. Unsurprisingly, our experiments show that the best strategy is problem dependent, but in many cases a cheaper strategy is at least as good as more expensive alternatives.

CCS CONCEPTS

•Computing methodologies → Gaussian processes; Modeling methodologies; •Applied computing → Multi-criterion optimization and decision-making; •Mathematics of computing → Probabilistic algorithms;

KEYWORDS

Computationally Expensive Optimisation; Efficient Multi-Objective Optimisation; Infill Criteria; Scalarisation methods.

ACM Reference format:
Alma A. M. Rahat, Richard M. Everson, and Jonathan E. Fieldsend. 2017. Alternative Infill Strategies for Expensive Multi-Objective Optimisation. In *Proceedings of GECCO '17, Berlin, Germany, July 15-19, 2017,* 8 pages.
DOI: http://dx.doi.org/10.1145/3071178.3071276

*Corresponding author

GECCO '17, Berlin, Germany

DOI: http://dx.doi.org/10.1145/3071178.3071276

1 INTRODUCTION

Real world multi-objective optimisation problems often consist of computationally or financially expensive objective functions. For instance, design optimisation of mechanical parts may require inspecting the performance of a design within a fluid environment using computational fluid dynamics (CFD) simulations. A high quality CFD simulation may take hours to converge, and thus only a limited number of designs may be considered in optimisation.

Many effective algorithms have been proposed in the last decade for expensive multi-objective optimisation, see for example [4, 6, 8, 20, 24]. Generally, these are model-based approaches inspired by single objective Bayesian global optimisation methods. Based on an initial set of expensively evaluated solutions, a Bayesian surrogate model, either for each objective (multi-surrogate) or for a scalarised representation of the multi-objective problem (mono-surrogate), is constructed. Regardless of what was modelled, a surrogate based multi-objective quality indicator, often referred to as an *infill criterion*, is derived. It is usually much cheaper to evaluate in comparison to the original objective functions, but frequently induces a highly multi-modal single objective fitness landscape. As such evolutionary optimisers perform well in locating promising solutions using the infill landscape. A candidate solution is then expensively evaluated, and the surrogate model(s) are retrained. The process is repeated until the budget on the expensive function evaluations is exhausted. Thus, these methods require only a few hundreds of expensive function evaluations to generate a good approximation of the optimal trade-off between multiple objectives.

One of the major issues with the most effective multi-surrogate infill criterion is that it often requires multi-dimensional integration, and therefore optimising it may become impractically expensive. A promising cheaper alternative are mono-surrogate approaches, but the only example of such an approach in a Bayesian optimisation framework is ParEGO [20]. Addressing these issues, the major contributions of this paper are as follows.

- We devise a novel infill criterion based on the minimum probability of improvement over an estimated Pareto set as an alternative multi-surrogate approach.
- We propose a range of set-based scalarisation functions modelling hypervolume improvement, dominance ranking or minimum signed distance from an estimated Pareto set, that may be used in a mono-surrogate Bayesian framework, and therefore promote research on this front.

The rest of the paper is structured as follows. In Section 2, we present the required background and the relevant work in the literature. The novel infill strategies are described in Sections 3 and 4. We present our results in Section 5. General conclusions are drawn in Section 6.

2 BACKGROUND

We now present a synthesis of the relevant background material.

2.1 Single Objective Efficient Global Optimisation (EGO)

Efficient Global Optimisation (EGO) or *Bayesian Optimisation (BO)* is a particular area of surrogate-assisted (evolutionary) optimisation. In practice, it has proved to be a very effective approach for single objective expensive optimisation problems with limited budget on the number of true function evaluations. A recent review on the topic can be found in [26].

EGO is essentially a global search strategy that sequentially samples the design space at likely locations of the global optimum [17]. It starts with a space filling design (e.g. Latin hypercube sampling [22]) of the parameter space, constructed independent of the function space. The solutions from this initial design are then evaluated with the true function. Using the set of the initial design parameters and the associated function values as data a regression model is trained. Promising parameters at which to evaluate the function can then be located using the surrogate. Frequently the surrogate model is a stochastic process, usually a Gaussian process ($\mathcal{GP}$)[1]. The benefit of using $\mathcal{GP}$s for regression is that they provide a posterior predictive distribution given the training data, and thus querying the surrogate model at any solution in the design space results in both a mean prediction and the uncertainty associated with the prediction. This often enables the closed form calculation of an *infill criterion*, that is the expected improvement in function value (with respect to the best function value observed so far) to be obtained by querying a solution. This infill criterion has monotonicity properties: it is inversely proportional to the predicted mean (with fixed uncertainty), and directly proportional to the uncertainty in prediction (with fixed predicted mean). As a consequence, it strikes a balance between global exploration and myopic exploitation of the model. Therefore, a strategy for selecting the next solution is to (expensively) evaluate the parameters that maximise the infill criterion. The newly sampled data is then added to the training database, and a retraining of the $\mathcal{GP}$ model ensues. The process is repeated until the budget is exhausted.

A single objective optimisation problem may be expressed as:

$$\min_{\mathbf{x}} f(\mathbf{x}), \tag{1}$$

where the parameters $\mathbf{x} \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$. With the initial design $\mathcal{D} = \{(\mathbf{x}^m, f^m = f(\mathbf{x}^m)\}_{m=1}^M$ of M samples, a $\mathcal{GP}$ model may be constructed. In essence, a $\mathcal{GP}$ is a collection of random variables, and any finite number of these have a joint Gaussian distribution [25]. The predictive density of the function for parameters $\mathbf{x}$ given by a $\mathcal{GP}$ model based on the observations $\mathcal{D}$ may be expressed as:

$$P(\hat{f}(\mathbf{x}) \mid \mathbf{x}, \mathcal{D}, \theta) = \mathcal{N}(\hat{f}(\mathbf{x}) \mid \mu(\mathbf{x}), \sigma^2(\mathbf{x})), \tag{2}$$

where the mean and variance are

$$\mu(\mathbf{x}) = \boldsymbol{\kappa}(\mathbf{x}, X) - K^{-1}\mathbf{f} \tag{3}$$

$$\sigma^2(\mathbf{x}) = \kappa(\mathbf{x}, \mathbf{x}) - \boldsymbol{\kappa}(\mathbf{x}, X)^\top K^{-1} \kappa(X, \mathbf{x}). \tag{4}$$

Here $X \in \mathbb{R}^{M \times n}$ is the matrix of observed parameter values and $\mathbf{f} \in \mathbb{R}^M$ is the corresponding vector of the true function evaluations; thus $\mathcal{D} = \{(X, \mathbf{f})\}$. The covariance matrix $K \in \mathbb{R}^{M \times M}$ represents the covariance function $\kappa(\mathbf{x}, \mathbf{x}')$ evaluated for each pair of observations and $\boldsymbol{\kappa}(\mathbf{x}, X) \in \mathbb{R}^M$ is the vector of covariances between $\mathbf{x}$ and each of the observations. In this paper, we use a flexible class of covariance functions embodied in the Matern 5/2 kernel, as recommended for modelling realistic functions [27]. We used the limited memory BFGS algorithm with 10 restarts to optimise the kernel hyperparameters; see [12] for details.

The predicted improvement over the best evaluated solution so far, $f^* = \min_m\{f^m(\mathbf{x}_m)\}$, is: $I(\mathbf{x}, f^*) = \max\ (f^* - \hat{f}(\mathbf{x}), 0)$. Therefore, the infill criterion (i.e. expected improvement at $\mathbf{x}$) based on the surrogate model may be expressed as [17]:

$$\alpha(\mathbf{x}, f^*) = \int_{-\infty}^{\infty} I(\mathbf{x}, f^*) P(\hat{f} \mid \mathbf{x}, \mathcal{D})\, d\hat{f} = \sigma(\mathbf{x})\, (s\Phi(s) - \phi(s)), \tag{5}$$

where $s = (f^* - \mu(\mathbf{x}))/\sigma(\mathbf{x})$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the Gaussian probability density function and cumulative density functions. The infill criterion is essentially the improvement weighted by the part of the posterior predictive distribution that lies below the evaluated minimum f^* and thus balances the exploitation of solutions which are very likely to be a little better than f^* with the exploration of others which may, with lower probability, turn out to be much better. Thus maximising this infill criterion estimates where the global optimum may be given the data and a strategy for determining the next solution to evaluate is the following maximisation problem:

$$\mathbf{x}^{M+1} = \underset{\mathbf{x}}{\operatorname{argmax}}\ \alpha(\mathbf{x}, \hat{f}). \tag{6}$$

The evaluation of the infill criterion in (5) is generally cheap. Thus an evolutionary algorithms may be used to locate an approximation of the optimal solution. This new solution may then be evaluated with the expensive function $f^{M+1} = f(\mathbf{x}^{M+1})$, and the dataset is augmented with the new solution $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\mathbf{x}^{M+1}, f^{M+1})\}$. The $\mathcal{GP}$ is retrained with the augmented dataset $\mathcal{D}$. The process is repeated until the limit on the number of expensive function evaluations is reached.

Note that the infill criterion may induce a highly multi-modal fitness landscape. Therefore locating the solution that maximises the expected improvement may require a large number of evaluations on the surrogate $\mathcal{GP}$ model. Hence, selecting the next solution to evaluate may be relatively expensive despite the fact that computing the expected improvement for a single $\mathbf{x}$ is cheap.

2.2 Multi-Objective Optimisation Problem

Many real world problems have multiple, often conflicting, objectives, and it is important to extremise these objectives simultaneously [5]. Consider a decision vector $\mathbf{x} \in \mathbb{R}^n$ within the feasible parameter space $\mathcal{X}$. Without loss of generality, a multi-objective optimisation problem with D objectives may then be expressed as:

$$\min_{\mathbf{x}} \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_D(\mathbf{x})), \tag{7}$$

where $f_i(\mathbf{x})$ is the ith objective and $\mathbf{F} : \mathcal{X} \in \mathbb{R}^n \to \mathbb{R}^D$ generates the objective space.

Due to the potentially conflicting objectives, generally there is not a unique solution to the optimisation problem, but a range of solutions trading-off between the objectives. The trade-off between solutions is characterised by the notion of dominance: a solution $\mathbf{x}$

[1] Gaussian Processes subsume Kriging.

is said to dominate another solution $\mathbf{x}'$, denoted as $\mathbf{x} \prec \mathbf{x}'$, iff

$$f_i(\mathbf{x}) \leq f_i(\mathbf{x}') \; \forall i = 1, \ldots, D \text{ and } f_i(\mathbf{x}) < f_i(\mathbf{x}') \text{ for some } i. \quad (8)$$

The set of solutions representing the optimal trade-off between the objectives is referred to as the Pareto set:

$$\mathcal{P} = \{\mathbf{x} \,|\, \mathbf{x}' \nprec \mathbf{x} \; \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X} \wedge \mathbf{x} \neq \mathbf{x}'\}, \quad (9)$$

and the image of the Pareto set in the objective space is known as the Pareto front $\mathcal{F} = \{\mathbf{F}(\mathbf{x}) \,|\, \mathbf{x} \in \mathcal{P}\}$.

Exactly locating the complete Pareto set may not be possible within a practical time limit, even for cheap objective functions, and an approximation is often sufficient. Therefore, the overall goal of an effective optimisation approach is to generate a good approximation of the Pareto set $\mathcal{P}^* \subseteq \mathcal{X}$.

Existing infill strategies for multi-objective optimisation based on $\mathcal{GP}$s may be categorised in two groups: multi-surrogate and mono-surrogate approaches. In multi-surrogate approaches, each objective function $f_i(\mathbf{x})$ is modelled. These models are often considered to be independent ignoring any potential cross-correlations between models, which is known to reduce overall uncertainty in predictions [19]. The combined models induce a multivariate Gaussian predictive distribution, with a diagonal covariance matrix: $p(\hat{\mathbf{F}}(\mathbf{x}) \,|\, \mathcal{D}) = \prod_i^D p(\hat{f}_i(\mathbf{x}) \,|\, \mathbf{x}, \mathcal{D}) = \mathcal{N}(\hat{\mathbf{F}} \,|\, \boldsymbol{\mu}(\mathbf{x}), \Sigma(\mathbf{x}))$, with $\boldsymbol{\mu}(\mathbf{x}) = (\mu_1(\mathbf{x}), \ldots, \mu_D(\mathbf{x}))$ and $\Sigma(\mathbf{x}) = \text{diag}(\sigma_1^2(\mathbf{x}), \ldots, \sigma_D^2(\mathbf{x}))$, from which an infill criterion is $\alpha(\mathbf{x}, \hat{\mathbf{F}})$ may be derived. On the other hand, mono-surrogate approaches aggregate the D objective functions to generate a scalarised model $g(\mathbf{x}) \equiv g(\mathbf{F}(\mathbf{x}))$. A surrogate model $\hat{g}$ of the scalarisation is used to compute the infill criterion $\alpha(\mathbf{x}, \hat{g})$. In both cases, the next solution to evaluate is the one which extremises the relevant infill criterion $\alpha(\cdot)$. Algorithm 1 briefly describes these efficient multi-objective optimisation approaches.

Algorithm 1 Efficient multi-objective optimisation.

Inputs

M : Number of initial samples
T : Budget on expensive function evaluations

Steps

1: $X \leftarrow \text{LatinHypercubeSampling}(\mathcal{X})$ ▷ Generate initial samples
2: $\mathbf{f} \leftarrow \mathbf{F}(\mathbf{x} \in X)$ ▷ Expensively evaluate all initial samples
3: **for** $i = M \to T$ **do**
4: **if** MultiSurrogate **then** ▷ Multi-surrogate approach
5: $\hat{\mathbf{F}} \leftarrow \text{Train}\mathcal{GP}\ (X, \mathbf{f})$ ▷ Train a model for each objective
6: $\mathbf{x}^* \leftarrow \text{argmax}_{\mathbf{x}}\, \alpha(\mathbf{x}, \mathcal{F}^*)$ ▷ Optimise infill criterion
7: **else** ▷ Mono-surrogate approach
8: $\hat{g} \leftarrow \text{Train}\mathcal{GP}\ (X, g(\mathbf{x} \in X))$ ▷ Train a model for scalarised objective
9: $\mathbf{x}^* \leftarrow \text{argmax}_{\mathbf{x}}\, \alpha(\mathbf{x}, \hat{g})$ ▷ Optimise infill criterion
10: **end if**
11: $X \leftarrow X \cup \{\mathbf{x}^*\}$ ▷ Augment data set with $\mathbf{x}^*$
12: $\mathbf{f} \leftarrow \mathbf{f} \cup \{\mathbf{F}(\mathbf{x}^*)\}$ ▷ Expensively evaluate $\mathbf{x}^*$
13: $\mathcal{P}^* \leftarrow \text{nondom}(X)$ ▷ Update Pareto set
14: $\mathcal{F}^* \leftarrow \text{nondom}(\mathbf{f})$ ▷ Update Pareto front
15: **end for**
16: **return** $\mathcal{P}^*$

Clearly, the infill criterion is a form of scalarisation of the original multi-objective problem. The central distinction between multi- and mono-surrogate approaches is therefore how this scalarisation is performed. In multi-surrogates, this scalarisation is based on predictive models, but in mono-surrogates the scalarisation is based on the deterministic evaluations of $\mathbf{F}$ and uncertainty in the prediction enters through a predictive model for the scalarisation.

2.3 Related Work

Most effective multi-surrogate strategies use expected hypervolume improvement as a multi-objective infill criterion. First proposed by Emmerich [8], the expected hypervolume improvement calculates the potential gain that may be achieved over the current Pareto set $\mathcal{P}^*$ by augmenting $\mathcal{P}^*$ with a solution based on its predictive distribution. This, however, involves multidimensional integration over the non-dominated objective space, which is achieved by decomposing the integration volume into disjoint cells and accounting for the volume weighted by the predictive distribution in each cell. As such the run time complexity is high and dependent on the number of solutions $|\mathcal{P}^*|$. Practical improvements on implementations of the expected hypervolume computation have been proposed by Hupkens *et al.* [16] and Couckuyt *et al.* [6], but the worst case time complexity is $O(|\mathcal{P}^*|^D)$ for $D = 2, 3$ objectives; for more objectives, the time complexity is conjectured to be even higher [16].

An alternative approach with a proxy for expected hypervolume improvement, referred to as $\mathcal{S}$-metric selection EGO (*SMS-EGO*), was proposed by Ponweiser *et al.* [24] and later improved by Wagner *et al.* [28]. In this approach, the posterior predictive distribution is accounted for implicitly with and overestimated mean prediction by simply subtracting the scaled uncertainty. This permits a deterministic calculation of hypervolume improvement over $\mathcal{P}^*$ for a tentative solution. Although this is comparatively cheaper to calculate, it still is expensive as the hypervolume calculation must be performed to evaluate the infill criterion for each tentative solution. Nonetheless, it has been shown to perform better or at least as well as the other methods [28]. We therefore choose to compare against SMS-EGO in this paper.

Other multi-surrogate infill strategies consider probability of improvement of a solution over $\mathcal{P}^*$ [6, 18], minimum Euclidean distance of mean predictions over $\mathcal{P}^*$ [18], aggregating the posterior prediction with Chebyshev scalarisation and computing the expected improvement in each scalarisation function within the MOEA/D framework [29], minimum angle penalised distance or maximum uncertainty within the reference vector guided evolutionary (RVEA) framework [4], etc.

The only mono-surrogate approach used within the Bayesian EGO framework is ParEGO [20]. It uses the normalised objective function values with an augmented Chebyshev function and a predefined set of weight vectors to achieve a scalarisation of the original multi-objective problem. The scalarised function is learned using a $\mathcal{GP}$ model and the standard expected improvement is calculated using equation (5). This mono-surrogate approach is known to be the considerably faster than other methods [4]. This is because only one model is maintained and trained (step 8 in Algorithm 1), and locating a solution that maximises the expected improvement is cheap because evaluation of the $\mathcal{GP}$ is inexpensive.

Therefore more research should be carried out in mono-surrogate approaches; especially given the success of set-based quality indicators in standard multi-objective evolutionary approaches, see for example IBEA [31] and HypE [2]. One of the main contributions of this paper is to propose a range of mono-surrogate strategies and thus propel research on this front. We therefore compare our infill criteria with ParEGO as well. Note that other mono-surrogate approaches, such as model based strategies proposed by Loschilov *et al.* [21] and Azzouz *et al.* [1], do not use $\mathcal{GP}$s, and thus may not be used within the Bayesian EGO framework.

3 MULTI-SURROGATE APPROACH: MINIMUM PROBABILITY OF IMPROVEMENT (MPOI)

In a multi-surrogate approach, we consider independent $\mathcal{GP}$ models for each objective: $p(\hat{f}_i(\mathbf{x}) \mid \mathbf{x}, \mathcal{D}) = \mathcal{N}(\hat{f}(\mathbf{x}) \mid \mu_i(\mathbf{x}), \sigma_i^2(\mathbf{x}))$. Assuming that the objectives are independent, the probability that a solution $\mathbf{x}$ dominates another solution $\mathbf{x}'$ is given by [9, 15]:

$$P(\mathbf{x} \prec \mathbf{x}') = \prod_{i=1}^{D} P(\hat{f}_i(\mathbf{x}) < \hat{f}_i(\mathbf{x}')), \tag{10}$$

where

$$P(\hat{f}_i(\mathbf{x}) < \hat{f}_i(\mathbf{x}')) = \frac{1}{2}\left[1 + \mathrm{erf}\left(m_i(\mathbf{x}, \mathbf{x}')/\sqrt{2}\right)\right], \tag{11}$$

$$\text{and } \; m_i(\mathbf{x}, \mathbf{x}') = \frac{\mu_i(\mathbf{x}') - \mu_i(\mathbf{x})}{\sqrt{\sigma_i^2(\mathbf{x}) + \sigma_i^2(\mathbf{x}')}}. \tag{12}$$

Note that since we consider the true evaluations to be noise-free, for any $\mathbf{x} \in X$, $\mu_i(\mathbf{x}) = f_i(\mathbf{x})$ and $\sigma_i^2(\mathbf{x}) = 0$.

Comparing an arbitrary solution $\mathbf{x}' \in \mathcal{X}$ with a solution $\mathbf{x} \in \mathcal{P}^*$, the current estimated Pareto set, there are three mutually exclusive possibilities: $\mathbf{x}'$ dominates $\mathbf{x}$ ($\mathbf{x}' \prec \mathbf{x}$), $\mathbf{x}'$ is dominated by $\mathbf{x}$ ($\mathbf{x} \prec \mathbf{x}'$), or they are mutually non-dominated ($\mathbf{x} \parallel \mathbf{x}'$). Therefore, the probability that $\mathbf{x}'$ improves upon a solution $\mathbf{x} \in \mathcal{P}^*$ is:

$$P(\mathbf{x}' \prec \mathbf{x} \text{ or } \mathbf{x} \parallel \mathbf{x}') = 1 - P(\mathbf{x} \prec \mathbf{x}'). \tag{13}$$

Intuitively, this measures the probability mass in the objective space beyond a solution $\mathbf{x} \in \mathcal{P}^*$, i.e. the space not dominated by $\mathbf{x}$, due to the multivariate predictive distribution for $\hat{\mathbf{F}}$. With this notion of the probability of improvement, we can define a multi-objective infill criterion based on least improvement upon any solution from the current Pareto front $\mathcal{F}^*$ of evaluated solutions:

$$\alpha_p(\mathbf{x}', \mathcal{F}^*) = \min_{\mathbf{x} \in \mathcal{P}^*} (1 - P(\mathbf{x} \prec \mathbf{x}')). \tag{14}$$

Thus, in a multi-objective EGO, the next solution to evaluate is:

$$\mathbf{x}^{M+1} = \underset{\mathbf{x}' \in \mathcal{X}}{\mathrm{argmax}} \; \alpha_p(\mathbf{x}', \mathcal{F}^*). \tag{15}$$

Keane [18] and Couckuyt *et al.* [6] suggested computing the probability of improvement over all solutions $\mathbf{x} \in \mathcal{P}^*$, i.e. $1 - P(\bigcup_{\mathbf{x} \in \mathcal{P}^*} \mathbf{x} \prec \mathbf{x}')$, as an infill criterion. Again, similar to expected hypervolume improvement calculations, this requires multi-dimensional integration by decomposing the non-dominated objective space into disjoint regions. As such, it is computationally expensive, especially for many-objective problems. In our formulation of the infill criterion $\alpha_p(\mathbf{x}', \mathcal{F}^*)$ in equation (14), we implicitly cover all the solutions from the current estimated Pareto set $\mathcal{P}^*$ by considering the minimum probability of improvement over $\mathcal{P}^*$. It is therefore fast to calculate, with the most expensive step being the calculation of $\mathrm{erf}(\cdot)$ function.

3.1 Monotonicity Properties

The role of an infill criterion is to help us choose a good candidate solution. The efficacy of an infill criterion thus depends on how well it distinguishes between two tentative solutions $\mathbf{x}', \mathbf{x}'' \in \mathcal{X} \setminus X$. A sound multi-objective infill criterion must therefore satisfy the following necessary conditions for two solutions $\mathbf{x}'$ and $\mathbf{x}''$ [28].

- N1 The dominance relationship should be preserved given that the uncertainty is equal. That is, if $\mu_i(\mathbf{x}') < \mu_i(\mathbf{x}'') \wedge \sigma_i(\mathbf{x}') = \sigma_i(\mathbf{x}''), \forall i \in \{1, \ldots, D\}$, then: $\alpha_p(\mathbf{x}', \mathcal{F}^*) > \alpha_p(\mathbf{x}'', \mathcal{F}^*)$.
- N2 When mean predictions are equal, the infill criterion should monotonically increase with the uncertainty. That is if $\mu_i(\mathbf{x}') = \mu_i(\mathbf{x}'') \wedge \sigma_i(\mathbf{x}') > \sigma_i(\mathbf{x}''), \forall i \in \{1, \ldots, D\}$, then: $\alpha_p(\mathbf{x}', \mathcal{F}^*) > \alpha_p(\mathbf{x}'', \mathcal{F}^*)$.

Proof. Clearly, from equation (14), it is sufficient to prove that for any solution $\mathbf{x} \in \mathcal{P}^*$, $P(\mathbf{x} \prec \mathbf{x}') < P(\mathbf{x} \prec \mathbf{x}'')$, for both N1 and N2 to be true. This further implies: it is equivalent to prove that $m_i(\mathbf{x}, \mathbf{x}') < m_i(\mathbf{x}, \mathbf{x}''), \forall i \in \{1, \ldots, D\}$ under the given conditions (c.f. equations (10) and (11)). As discussed earlier, with no noise in measurements $\mu_i(\mathbf{x}) = f_i(\mathbf{x})$ and $\sigma_i(\mathbf{x}) = 0$. Now, considering equation (12), given the same uncertainty in prediction $\sigma_i(\mathbf{x}') = \sigma_i(\mathbf{x}'')$, $m_i(\mathbf{x}, \mathbf{x}') < m_i(\mathbf{x}, \mathbf{x}'')$ for ith objective iff $\mu_i(\mathbf{x}') < \mu_i(\mathbf{x}'')$. Thus N1 is satisfied. Similarly, when the mean predictions are the same $\mu_i(\mathbf{x}') = \mu_i(\mathbf{x}'')$, then $m_i(\mathbf{x}, \mathbf{x}') < m_i(\mathbf{x}, \mathbf{x}'')$ for ith objective iff $\sigma_i(\mathbf{x}') > \sigma_i(\mathbf{x}'')$. Therefore N2 is satisfied. ■

3.2 Infill Fitness Landscape

An infill criterion essentially is a scalar representation of the objective space. It is therefore interesting to investigate the induced infill landscape within the objective space.

To achieve a visual impression of the landscape, we consider evenly distributed samples from the objective space. Considering each sample as the mean prediction of a multivariate $\mathcal{GP}$, and setting a fixed uncertainty $\sigma = 0.1$, we calculate the minimum probability of improvement. In Figure 1, we show the resulting characterisation of the objective space. We can observe a clear indication that optimising this infill criteria promotes sampling in the non-dominated region, and hence it is likely to improve the current estimation of the Pareto set. It is also evident that the single objective infill criteria is highly multi-modal. It should be noted that in reality the characterisation using trained models may appear different due to variations in uncertainty. Nonetheless, the infill criteria adheres to the monotonicity properties as described in Section 3.1.

4 MONO-SURROGATE APPROACH: INDICATOR BASED SCALARISATION

To the best of our knowledge the only scalarising function used in a mono-surrogate framework is an augmented Chebyshev function [20]. However, it is possible to model other set-based quality indicators that measure the rank of a solution within the expensively sampled solutions instead. Such scalarised ranking can then

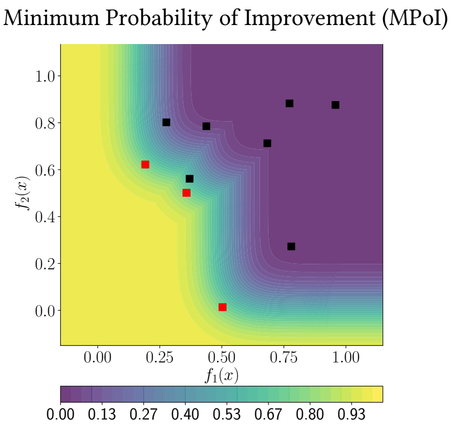

Figure 1: Infill fitness landscape in a hypothetical two-objective space with fixed uncertainty $\sigma = 0.1$ in prediction for minimum probability of improvement over current estimated Pareto front $\mathcal{F}^*$ (red squares). Lighter shades depict higher probability of improvement, and thus these areas are preferred over darker areas. Maximising the criterion promotes sampling in the non-dominated objective space. The black squares show the dominated solutions in the data set. The contours are unaffected by the dominated solutions.

be modelled with a $\mathcal{GP}$ and used within the EGO framework. So long as a scalarisation function preserves dominance relationship, maximising such scalarisation should improve the current set [31]. Although many different scalarisation methods may be used, in this section we present three scalarisation methods.

4.1 Hypervolume Improvement (HypI)

The hypervolume is a set-based quality indicator that measures the objective space covered between a non-dominated set and a predefined reference vector [30]. It is an exceptional set-based indicator as Fleischer proved that maximising hypervolume is equivalent to locating the optimal Pareto set [10]. We therefore propose a scalarisation based on hypervolume improvement.

A set of expensively sampled solutions $\mathcal{D} = \{(X, \mathbf{f} = F(X))\}$ can be ranked according to the "Pareto shell" in which they lie. Let the first Pareto shell be the estimated Pareto set for X: $\mathcal{P}'_1 = \mathcal{P}^* = \text{nondom}(X)$, where the nondom$(\cdot)$ function returns the maximal non-dominated subset of its argument. Then successive Pareto shells $\mathcal{P}'_l$ $(l > 1)$ are defined as:

$$\mathcal{P}'_l = \text{nondom}(X \setminus \cup_{i<l}\mathcal{P}_i). \tag{16}$$

The hypervolume indicator for a set X is the volume of objective space which is dominated by solutions in X and which dominates a reference vector $\mathbf{r}$ [30]:

$$H(X, \mathbf{r}) = \text{vol}_{\mathbf{z}}\{\mathbf{F}(\mathbf{x}) \prec \mathbf{z} \prec \mathbf{r} \wedge \mathbf{x} \in X\}. \tag{17}$$

In order to define the hypervolume improvement due to a solution $\mathbf{x}$, we consider the first Pareto shell that contains no solutions that dominate $\mathbf{x}$; denote this shell by $\mathcal{P}'_k$. Then the hypervolume improvement for $\mathbf{x}$ is the hypervolume corresponding to $\mathcal{P}'_k$ augmented with $\mathbf{x}$:

$$g_h(\mathbf{x}, X) = H(\mathbf{x} \cup \mathcal{P}'_k, r). \tag{18}$$

Here the function $g_h(\mathbf{x}, X)$ generates a set-based scalarisation of the original multi-objective problem for the set of sampled solutions. Clearly, if a solution $\mathbf{x}''$ is dominated by another $\mathbf{x}'$, then $g_h(\mathbf{x}', X) > g_h(\mathbf{x}'', X)$. Therefore we may learn a function $\hat{g}_h(\mathbf{x})$ and use the expected improvement in equation (5) within EGO framework in order to maximise the hypervolume improvement (a *unary* indicator), and thus improve upon the current approximation of the Pareto set $\mathcal{P}^*$.

Note that this set-based scalarisation is different from the formulation HypE [2]. In HypE, the scalarisation does not differentiate between the dominated solutions during selection, and hence all dominated solutions have the same scalar value associated with the,, creating a plateau of equal fitness behind $\mathcal{P}^*$. Using a $\mathcal{GP}$ to model the scalarisation thus has limited spatial information, which may hinder EGO's global exploration. In contrast, we base our scalarisation purely on dominance rank based hypervolume improvement, and in order to generate a positive fitness gradient towards the Pareto set, aiding the EGO process.

4.2 Dominance Ranking (DomRank)

In MOGA [11], a fitness assignment scheme based on dominance was proposed. In essence, a solution is assigned a rank in proportion to the number of already-evaluated solutions that dominate it. We use a straightforward adaptation of this.

Given a set of solutions X, a solution can be dominated by at most $|X| - 1$ others. The ranking strategy is defined as:

$$g_c(\mathbf{x}, X) = 1 - \frac{|\{\mathbf{x}' | \mathbf{x}' \prec \mathbf{x} \wedge \mathbf{x} \neq \mathbf{x}', \forall \mathbf{x}, \mathbf{x}' \in X\}|}{|X| - 1}. \tag{19}$$

Therefore, the current estimated Pareto set members $\mathbf{x}^a \in \mathcal{P}^*$ have the maximum rank $g_c(\mathbf{x}^a, X) = 1$. Similarly, a solution $\mathbf{x}^b$ that is dominated by all other members is assigned rank $g_c(\mathbf{x}^b, X) = 0$.

By definition it is dominance preserving: if $\mathbf{x}' \prec \mathbf{x}''$, then $g_c(\mathbf{x}', X) > g_c(\mathbf{x}'', X)$. It is therefore suitable for use in the EGO framework.

4.3 Minimum Signed Distance (MSD)

Distance of a solution from the current estimate of the Pareto front is clearly important. Here, we consider the minimum signed distance of a solution from $\mathcal{P}^*$ as a ranking strategy:

$$g_d(\mathbf{x}, X) = \min_{\mathbf{x}' \in \mathcal{P}^*} d(\mathbf{x}', \mathbf{x}), \tag{20}$$

where $d(\mathbf{x}', \mathbf{x}) = \sum_{i=1}^{D} f_i(\mathbf{x}') - f_i(\mathbf{x})$ is a signed distance. Similar to the other measures, it is also dominance preserving and therefore if $\mathbf{x}' \prec \mathbf{x}''$, then $g_d(\mathbf{x}', X) > g_d(\mathbf{x}'', X)$.

4.4 Scalarisation Fitness Landscape

The mono-surrogate strategies also induce a fitness landscape in the objective space, which can be informatively visualised. Rather than considering solutions from the decision space, we use evenly distributed points in the multi-dimensional objective space, and compute the fitness. In Figure 2, the resulting characterisations of the objective space is presented.

The hypervolume improvement (Figure 2 *left*) and the minimum signed distance (Figure 2 *right*) both show strong selection preference towards the minimum of both objectives. In comparison,

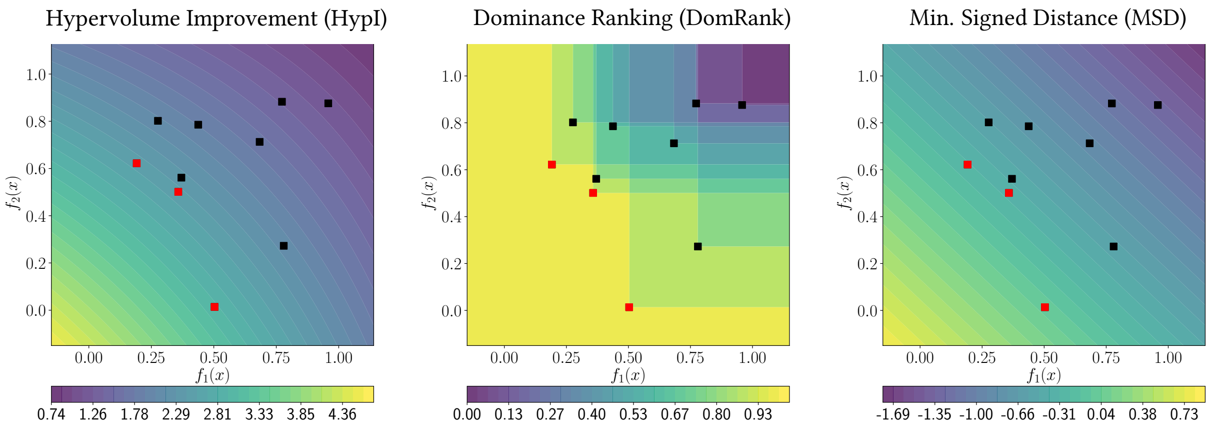

Figure 2: Scalarisation fitness landscape for hypervolume improvement (*left*; with reference vector $\mathbf{r} = (2, 2)$), dominance ranking (*middle*) and minimum signed distance (*right*) in a hypothetical two-objective space with respect to a data set shown in red (non-dominated) and black (dominated) squares. Lightly shaded areas represent higher fitness, so that maximising a scalarisation should lead to improving the current estimated Pareto set $\mathcal{P}^*$.

dominance ranking (Figure:2 *middle*) is proportional to the number of points that dominate a given point, and clearly does not discriminate between solutions in the non-dominated space. The figure demonstrates the dominance preserving property: maximising any of these measures should select solutions that improve $\mathcal{P}^*$.

5 ILLUSTRATION

To demonstrate the performance of the proposed strategies, we selected test problems of varying difficulty from the popular DTLZ [7] and WFG [14] problem suites. The selected test problems and relevant settings are summarised in Table 1.

Table 1: Selected test problems and relevant set up.

Problem	Parameters[2] n	Objectives D	Reference vector $\mathbf{r}$
DTLZ1	6	3	(400, 400, 400)
DTLZ2	6	3	(2.5, 2.5, 2.5)
DTLZ5	6	6	(2.5, 2.5, 2.5, 2.5, 2.5, 2.5)
DTLZ7	6	4	(1, 1, 1, 50)
WFG1	6	2	(10, 10)
WFG2	6	2	(10, 10)

The primary benefit of using these problems is that the Pareto set, and consequently the Pareto front, is known. Thus it allows us to either compute or approximate the optimal hypervolume [30] and can be used as a yardstick for competing methods. The Pareto set for DTLZ1 is a simplex with vertices at 0.5 in three dimensional objective space with many local fronts. For DTLZ2, the optimal front is a unit sphere in the positive octant. The shape of the Pareto front for DTLZ5 is unclear for four or more objectives [14]. In four dimensional objective space, DTLZ7 has a Pareto set that produces $2^{4-1} = 8$ disconnected Pareto-optimal regions. WFG1 has flat regions in the Pareto front and is strongly biased towards smaller values of the decision variables. The Pareto front for WFG2 consists of disconnected regions. Using the geometry of the DTLZ1 and DTLZ2 fronts, we calculated the optimal hypervolume. For the other problems, we approximated the optimal hypervolume with 10^4, 10^5 and 10^6 random members from the Pareto set for two, four and six objectives respectively.

We compare the proposed strategies with SMS-EGO (known for performance) [24, 28], ParEGO (known for speed) [20] and maximin Latin Hypercube Samples (LHS) with equal budget. Following the suggestions in [24], we implemented the dominance comparison in C for SMS-EGO. In ParEGO, we used 20 and 21 scalarising vectors for four and six objective problems respectively. The rest of the settings are standard for these algorithms, unless explicitly specified. It should be noted that SMS-EGO and ParEGO have not been tested in more than three objectives before in the literature.

In our experiments, we consider 11 runs of each method on any problem, starting from $11n - 1 = 65$ initial maximin LHS samples of the $n = 6$ dimensional decision space, and a budget of 250 function evaluations. These simulation runs are matched: different methods use the same initial design for a specific run and a specific problem; except for the competing independent LHS designs with 250 solutions. The performance of the strategies are investigated in terms of hypervolume of the estimated Pareto set.

The infill criterion landscape is usually highly multi-modal. Therefore, we used Bipop-CMA-ES [13]—which is known to perform well in solving multi-modal problems—to locate the solution that maximises an infill criterion. For a fair comparison, we used the same optimiser for SMS-EGO and ParEGO. To locate a good candidate solution, we set the the maximum number of infill criterion evaluation to $20000n$ based on the results from a short experiment in optimising any infill criteria. Otherwise, recommended settings for Bipop-CMA-ES were used.

We used statistical testing methods to determine which strategy performed best [23]. As we used matched initial samples, the Friedman test was performed to determine if there was a difference between all mono- and multi-surrogate approaches. Since we found a significant difference, a further multiple comparison test using the Wilcoxon Signed Rank test with Bonferroni correction was performed to identify the overall winner in a specific problem [3]. The comparisons between LHS and other methods were performed using the Mann-Whitney-U test [23], as the samples were

[2]The WFG problems are configured with two position and four distance parameters.

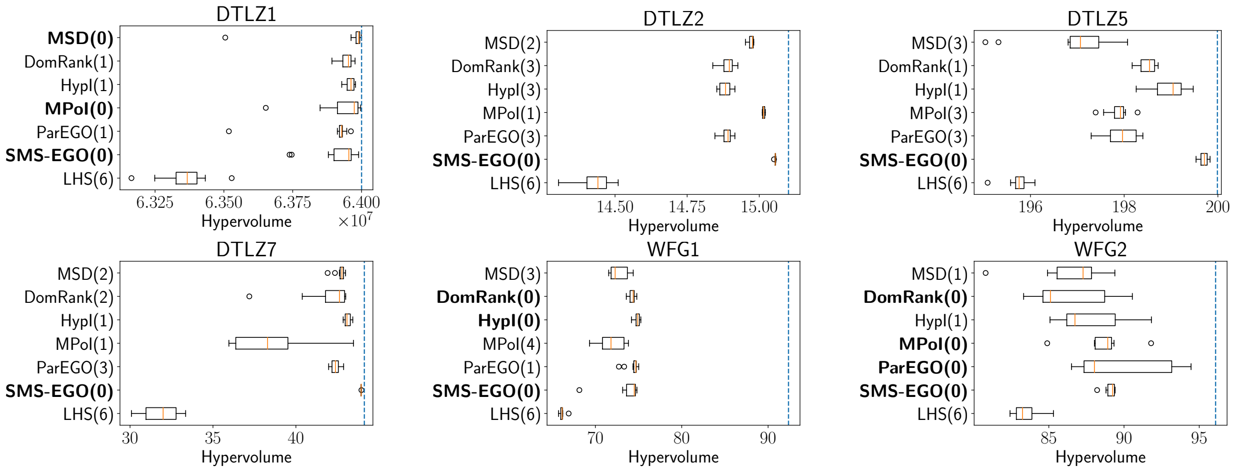

Figure 3: Hypervolume comparison for different strategies in selected test problems over 11 runs. Labels on the vertical axes show the method and the number of competing methods (in parentheses) that tested significantly better than the named strategy, so that lower number represent a better methods. The statistically best method(s) in a problem are highlighted with bold text. The optimal hypervolume is shown with blue dashed vertical lines.

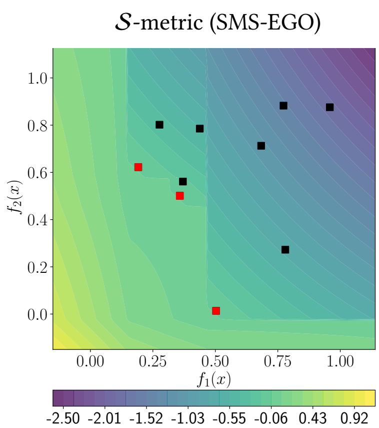

Figure 4: Infill fitness landscape in a hypothetical two-objective space with fixed uncertainty $\sigma = 0.1$ in prediction for $\mathcal{S}$-metric in SMS-EGO. The red squares depict the current estimated Pareto front $\mathcal{F}^*$ and the black squares show the dominated solutions in this data set.

not matched in this case. The maximum significance level was set to $\rho = 0.05$ in all cases. The results are presented in Figure 3.

The performance of the competing strategies are generally problem dependent. Based on the results from the statistical testing for performance comparison in all problems, the strategies considered here may be ranked in the following order: SMS-EGO, hypervolume improvement (HypI), dominance ranking (DomRank), minimum probability of improvement (MPoI), minimum signed distance (MSD), ParEGO, and Latin hypercube sampling (LHS). The simulation runs in Figure 3, indicate that all Bayesian methods are significantly better than naive LHS designs. Clearly, using a more informed data driven approach is better with limited budget on function evaluations.

SMS-EGO performs well across all problems, and it is never significantly worse than any other competing method. In particular, it outperforms other methods in DTLZ2, DTLZ5 and DTLZ7. However, in comparatively more difficult problems: DTLZ1, WFG1 and WFG2, which are difficult to model with a $\mathcal{GP}$ and a stationary kernel function, at least two of our proposed methods are competitive.

In contrast, ParEGO is significantly worse than at least one of the proposed methods, except on WFG2 where it is comparable to the best performing methods. In addition, ParEGO is also comparable to SMS-EGO in DTLZ1 and WFG1. However, in these two problems, MSD and HypI are better than ParEGO.

For simpler problems, we attribute the performance of a method to its characteristic fitness landscape. The $\mathcal{S}$-metric in SMS-EGO (Figure 4) is calculated considering optimistic model predictions, i.e. mean predictions scaled down with the associated uncertainties, and thus it extends the attainment surface towards the ideal objective vector. In practice this works well by focusing the search in the vicinity of the current front and in the non-dominated regions. In contrast, MPoI (Figure 1) and DomRank (Figure 2: *middle*) infill fitness contours closely follow the attainment surface. Consequently, these promote exploration of solutions at the edges due to the favourable mean predictions combined with large uncertainties far from observed solutions leading to high expected improvement. Although HypI (Figure 2: *left*) and MSD (Figure 2: *right*) do not show such bias towards the edges, these prefer the solution closest to the ideal vector and therefore the search may be somewhat misled.

For harder problems, the subtle differences in infill fitness landscape have little impact because of an imperfect model. As such, the methods presented here are mostly equivalent.

We also investigated the computation time for the infill criteria on Intel (i7-2.6GHz) machines.[3] It is clear from Figure 5 that SMS-EGO can be computationally expensive. The high computation cost in evaluating infill criterion is primarily due to hypervolume calculation, and consequently scales poorly with the number of objectives and the number of elements in the current estimated Pareto set. All other strategies are orders of magnitude faster than SMS-EGO, while the performance is comparable in many cases.

[3] Supplementary Figures and Python code for all the strategies used in this paper are available at: http://bitbucket.org/arahat/gecco-2017

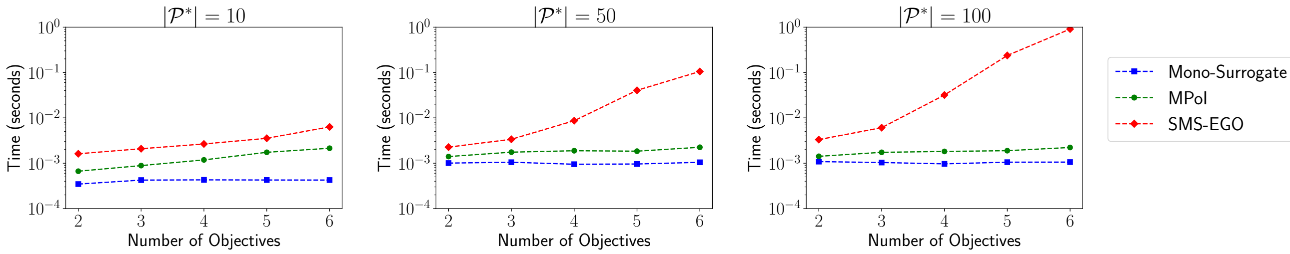

Figure 5: Comparison of average computation time per infill criterion evaluation over 1000 runs between strategies. SMS-EGO is highly dependent on the number of objectives and number of elements in the estimated Pareto set. Multi-surrogate MPoI is more expensive than mono-surrogate (including ParEGO) approaches. Both MPoI and mono-surrogate approaches are relatively insensitive to the increase in number of objectives or number of Pareto set elements, and orders of magnitude faster than SMS-EGO.

6 CONCLUSIONS

In this paper, we presented a novel cheaper multi-surrogate infill criterion based on minimum probability of improvement. We have also investigated a range of scalarisation functions modelling hypervolume improvement, dominance ranking or minimum signed distance from the estimated front. These effectively enable us to perform multi-objective Bayesian optimisation in a parameter free manner. The proposed fast infill strategies perform as well as SMS-EGO in half of the test problems presented here, while outperforming ParEGO. Current work focuses on the efficacy of various indicator functions and their ensembles within a mono-surrogate EGO framework.

ACKNOWLEDGMENTS

This research was supported by the Engineering and Physical Sciences Research Council [grant number EP/M017915/1].

REFERENCES

[1] N. Azzouz, S. Bechikh, and L. Ben Said. 2014. Steady state IBEA assisted by MLP neural networks for expensive multi-objective optimization problems. In *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation.* ACM, 581–588.

[2] J. Bader and E. Zitzler. 2011. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary computation* 19, 1 (2011), 45–76.

[3] R. Bender and S. Lange. 2001. Adjusting for multiple testing: when and how? *Journal of Clinical Epidemiology* 54, 4 (2001), 343 – 349.

[4] T. Chugh, Y. Jin, K. Miettinen, J. Hakanen, and K. Sindhya. 2016. A Surrogate-assisted Reference Vector Guided Evolutionary Algorithm for Computationally Expensive Many-objective Optimization. *IEEE Transactions on Evolutionary Computation* PP, 99 (2016), 1–14.

[5] C. A. C. Coello, G. B. Lamont, and D. A. Van Veldhuizen. 2007. *Evolutionary algorithms for solving multi-objective problems* (2nd ed.). Springer.

[6] I. Couckuyt, D. Deschrijver, and T. Dhaene. 2014. Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. *Journal of Global Optimization* 60, 3 (2014), 575–594.

[7] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. 2005. *Scalable test problems for evolutionary multiobjective optimization.* Springer.

[8] M. Emmerich. 2005. *Single-and multi-objective evolutionary design optimization assisted by Gaussian random field metamodels.* Ph.D. Dissertation.

[9] J. E. Fieldsend and R. M. Everson. 2005. Multi-objective optimisation in the presence of uncertainty. In *The 2005 IEEE Congress on Evolutionary Computation.*, Vol. 1. IEEE, 243–250.

[10] M. Fleischer. 2003. The measure of Pareto optima applications to multi-objective metaheuristics. In *International Conference on Evolutionary Multi-Criterion Optimization.* Springer, 519–533.

[11] C. M. Fonseca and P. J. Fleming. 1993. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization.. In *ICGA*, Vol. 93. 416–423.

[12] GPy. since 2012. GPy: A Gaussian process framework in Python. http://github.com/SheffieldML/GPy. (since 2012).

[13] N. Hansen. 2009. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers.* ACM, 2389–2396.

[14] S. Huband, P. Hingston, L. Barone, and L. While. 2006. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation* 10, 5 (2006), 477–506.

[15] E. J. Hughes. 2001. Evolutionary multi-objective ranking with uncertainty and noise. In *International Conference on Evolutionary Multi-Criterion Optimization.* Springer, 329–343.

[16] I. Hupkens, M. Emmerich, and A. Deutz. 2014. Faster computation of expected hypervolume improvement. *arXiv preprint arXiv:1408.7114* (2014).

[17] D. R. Jones, M. Schonlau, and W. J. Welch. 1998. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* 13, 4 (1998), 455–492.

[18] A. J. Keane. 2006. Statistical improvement criteria for use in multiobjective design optimization. *AIAA journal* 44, 4 (2006), 879–891.

[19] J. P. C. Kleijnen and E. Mehdad. 2014. Multivariate versus univariate Kriging metamodels for multi-response simulation models. *European Journal of Operational Research* 236, 2 (2014), 573–582.

[20] J. Knowles. 2006. ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation* 10, 1 (Feb 2006), 50–66.

[21] I. Loshchilov, M. Schoenauer, and M. Sebag. 2010. A mono surrogate for multiobjective optimization. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation.* ACM, 471–478.

[22] M. D. McKay, R. J. Beckman, and W. J. Conover. 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 42, 1 (2000), 55–61.

[23] J. D. Knowles nad L. Theile and E. Zitzler. 2006. *A Tutorial on the Performance Assesment of Stochastic Multiobjective Optimizers.* Technical Report TIK214. Computer Engineering and Networks Laboratory, ETH Zurich, Zurich, Switzerland.

[24] W. Ponweiser, T. Wagner, D. Biermann, and M. Vincze. 2008. Multiobjective optimization on a limited budget of evaluations using model-assisted $\mathcal{S}$-metric selection. In *International Conference on Parallel Problem Solving from Nature.* Springer, 784–794.

[25] C. E. Rasmussen and C. K. I. Williams. 2006. *Gaussian processes for machine learning.* The MIT Press.

[26] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. 2016. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* 104, 1 (2016), 148–175.

[27] J. Snoek, H. Larochelle, and R. P. Adams. 2012. Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems.* 2951–2959.

[28] T. Wagner, M. Emmerich, A. Deutz, and W. Ponweiser. 2010. On expected-improvement criteria for model-based multi-objective optimization. In *International Conference on Parallel Problem Solving from Nature.* Springer, 718–727.

[29] Q. Zhang, W. Liu, E. Tsang, and B. Virginas. 2010. Expensive Multiobjective Optimization by MOEA/D with Gaussian Process Model. *IEEE Transactions on Evolutionary Computation* 14, 3 (June 2010), 456–474.

[30] E. Zitzler. 1999. *Evolutionary algorithms for multiobjective optimization: Methods and applications.* Ph.D. Dissertation.

[31] E. Zitzler and S. Künzli. 2004. Indicator-based selection in multiobjective search. In *International Conference on Parallel Problem Solving from Nature.* Springer, 832–842.